Secondary Resource Allocation for Opportunistic Spectrum Sharing with IR-HARQ based Primary Users

Romain Tajan, Charly Poulliat, Inbar Fijalkow

Abstract

We propose to address the problem of a secondary resource allocation when a primary Incremental Redundancy Hybrid Automatic Repeat reQuest (IR-HARQ) protocol. The Secondary Users (SUs) intend to use their knowledge of the IR-HARQ protocol to maximize their long-term throughput under a constraint of minimal Primary Users (PUs) throughput. The ACcumulated Mutual Information (ACMI), required to model the primary IR-HARQ protocol, is used to define a Constrained Markov Decision Process (CMDP). The SUs resource allocation is then shown to be a solution of this CMDP. The allocation problem is then considered as an infinite dimensional space linear programming. Solving the dual of this linear programming is similar to solving an unconstrained MDP. A solution is finally given using the Relative Value Iteration (RVI) algorithm.

I. Introduction

Cognitive radio has been introduced in order to improve the efficiency of wireless networks (see e.g. [1], [2] and [3]). This paradigm allows Secondary Users (SUs) to opportunistically access the bandwidth of licensed Primary Users (PUs) adapting their parameters to limit their impact on the PUs performances. Initial works often consider an Opportunistic Spectrum Access (OSA) model for SUs. In this model, the SUs sense the PUs bandwidth intending to detect *white spaces* to communicate. OSA model targets a zero-interference policy. However, more recently an other model called Opportunistic Spectrum Sharing (OSS) has been proposed. OSS model

For this work R. Tajan was supported by DGA and CNRS

allows the SUs to interfere the PUs as long as the degradations on the PUs performances remain below a certain level.

Resource allocation has already been proven to be an efficient tool to address OSS problems. In [4] the authors propose a power allocation that maximizes the secondary ergodic capacity under peak power constraint and under average interference-power constraint. Their model takes in particular sensing imperfections into account. In [5], the authors propose different constrained power allocations. In particular, they consider a constraint of average interference-power, peak power or outage probability for primary and secondary users. These two papers consider "secondary centric" constraints. Indeed, they only consider constraints taking into account secondary parameters like secondary peak power or secondary average power. Even though an outage probability constraint is considered in [5], this constraint is mapped onto a secondary constraint.

In [6] the author compares capacity regions of two power allocations coming from two different optimization problems. The first allocation is the solution to the maximization of the SUs ergodic capacity under the constraints of peak power, average power and average interference power. The second allocation is the same as the first one except that the constraint on average interference power is replaced by a PUs ergodic capacity loss. The results shown in [6] leads to the conclusion that taking an "ergodic capacity loss" increases the ergodic capacity region. This idea is then used in [7] to show that it is worth considering the primary protocol while allocating secondary resources. Indeed, in [7] the secondary users realizes an *active learning* in order to get some insights on the instantaneous interference its creates on the primary. These insights are then processed to realize joint power and rate allocation.

In our case, we will consider that the primary does not exploit any Channel State Information at the Transmitter (CSIT), then it will not probe the channel to adapt its transmission parameters. This make useless the active learning proposed in [7]. However, when the primary user implements an ARQ or Hybrid-ARQ protocol, some learning can still be done by listening to the feedbacks required by the protocol. In [8], the authors study a cognitive channel where the PUs implement an ARQ protocol. They propose an information theoretical based approach for studying the secondary user capacity. The secondary user tracks the primary feedbacks which allow him to improve its throughput while limiting the primary throughput loss. In [9], a secondary power allocation has been studied when the primary user is implementing

Incremental Redundancy Hybrid-ARQ protocol. Unfortunately, the protocol proposed in [9] is restricted to HARQ with only two transmissions. Moreover, no optimization of rate or power has been proposed. This work has been extended to IR-HARQ with multiple rounds in [10] but still no power allocation is done by the secondary user. [11] proposes to use ARQ protocol of the primary user in order to manage the interferences generated by the secondary user. To do so, they propose a *finite* Constrained Markov Decision Process (CMDP) (see e.g. [12], [13], [14], [15]) to describe the state of the primary ARQ protocol. The secondary user communicates with fixed power and looks for the optimal on/off strategy. The choice of the action (on or off) is done in order to maximize the throughput of the secondary system under the constraint of primary throughput loss.

The main contributions of this paper are the following. We show that a model of CMDP can also be used to describe the primary IR-HARQ protocol. The main difference with the one proposed in [11] is that the CMDP model we propose is not finite. Indeed, our model is based on the evolution of ACcumulated Mutual Information (ACMI) (see e.g. [16], [17]). ACMI is used in order to analyse the long term throughput (average number of received bits per unit of time) of the IR-HARQ. Since the ACMI is a continuous random variable, we proposed a model of CMDP with Borel state and action spaces in order to realize a joint power and rate allocation for the secondary user. Finally, we derive an algorithm based on the Value Iteration to approximate a solution. The solution obtain in this paper is also different from the one of [18] since in this paper, no instantaneous CSIT is required at Tx_2.

This paper is organized as follows. In section II, a description of the considered network and of the primary and secondary protocols is done. In section III we present how the model can be associated with Constrained Markov Decision Process. In section IV, we show that there exists a solution to the proposed CMDP. In section V we show that the proposed CMDP can be viewed as a linear programming on infinite dimensional. In this section, we also give an algorithm based on the dual of the linear programming that that allows us to compute a solution. In section VI, we give some simulation results. The conclusion of this work is finally presented in section VII.

II. System description and performances

In this section we describe the network composed of the PUs and the SUs. We then present the HARQ protocol implemented by the PUs and the power and rate allocation used by SUs. To

study the two access protocols, we introduce the long-term throughput. The long-term throughput is a figure of merit employed to study the performances of HARQ protocols. We finally show how the OSS model can be solved as a constrained optimization problem.

A. System Model

We consider the network illustrated in Figure 1. It is composed of a primary and a secondary

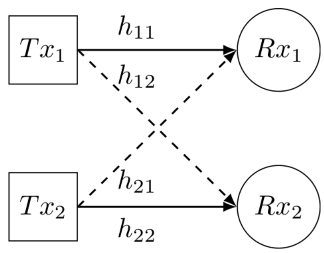

Figure 1. Model of the considered network

transmitters (respectively denoted Tx_1 and Tx_2) intending to send packets to their respective receivers Rx_1 and Rx_2. The transmission of a packet by Tx_1 incurs interferences on the received packet at Rx_2. And conversely, the transmission of a packet by Tx_2 incurs interferences on the received packet at Rx_1. Furthermore, we consider that the channel is a slotted block fading channel between Tx_i and Rx_j ($i, j \in \{1, 2\}$ where indexes 1 and 2 are respectively designating the PUs and the SUs). We assume stationary and ergodic channels such that the channel gains h_{ij}^n are assumed constant over all the duration of the slot n. For all the slots n, we also assume that h_{ij}^n is independent from $h_{i',j'}^n$ with $(i, j) \neq (i', j')$ that h_{ij}^n is independent from the noise and that h_{ij}^n is an independent and identically distributed (i.i.d) random variable such as $\alpha_{ij}^n = |h_{ij}^n|^2$ is an exponential random variable with mean $\overline{\alpha_{ij}}$. The signal received at receiver $i \in \{1, 2\}$ is then given as follows:

$$y_i^n = h_{ii}^n x_i^n + h_{ji}^n x_j^n + z_i^n, \tag{1}$$

where $z_i^n \in \mathbb{C}^L$ represents the noise which is a complex circular white Gaussian random vector of size L channel uses that we consider without loss of generality of zero mean and unit variance. The input signals $x_i^n \in \mathbb{C}^L$ and x_j^n are assumed to be complex circular white Gaussian random vectors of zero mean and respective transmit powers p_1^n and p_2^n. We further assume that the

$y_i^n \in \mathbb{C}^L$ are messages received at Rx_i. The instantaneous Signal to Interference plus Noise Ratio (SINR) at Rx_i at slot n is denoted by β_i^n and is defined as follows:

$$\beta_i^n \triangleq \frac{p_i \alpha_{ii}^n}{1 + p_j \alpha_{ji}^n}. \tag{2}$$

B. The primary protocol

We suppose that the primary user implements an IR-type HARQ protocol (see e.g. [16]). The transmitter Tx_1 encodes packet of size b_1 $bits$ denoted by u into a codeword denoted by x of length NL cu.The codeword x is then divided into N codeblocks namely $x_1, x_2, ..., x_N$ of length L cu. The protocol happens as follow: Tx_1 sends the codeword x_1 into the channel with equivalent rate obtained as $r_1 = \frac{b_1}{L} = Nr_1'$. If Rx_1 decodes x_1 successfully, it broadcasts an acknowledgement (ACK) bit into a feedback channel and Tx_1 starts the transmission of the next packet in its queue. We consider in this paper that Tx_1 is backlogged, i.e. we consider that Tx_1 always has a packet to transmit in its buffer. We further assume the one-bit feedback channel to be instantaneous and error-free. If the decoding of x_1 at Rx_1 fails, it sends a negative acknowledgement (NACK) bit in the channel. Tx_1 will then transmit x_2 on the next slot. When Rx_1 receives x_2, it does code combining between x_1 and x_2 and tries again to decode. This protocol keeps going until either Rx_1 successfully decodes the current information packet or the N codewords are used and the decoding of x fails. If the decoding is still unsuccessful after the N transmissions, an *outage* is declared and we assume that the packet is discarded.

We will finally assume that Tx_1 and Rx_1 are oblivious of the presence of the secondary users so that they do not modify their transmission parameters r_1 and p_1 dependently on the presence or not of the secondary user.

C. The secondary protocol

We suppose that the secondary users can listen to the primary feedbacks (see [8]) of the primary users and use these feedbacks to adjust an Adaptive modulation and Coding (AMC) scheme. Letting n be the index of the current slot, Tx_2 chooses its rate and power $(p_2^n, r_2^n) \in [0, P_{2M}] \times [0, R_{2M}]$ accordingly to the primary state. P_{2M} is the maximum peak power allowed for Tx_2 and the maximum rates R_{2M} in the AMC. We further assume that there is no Channel State Information at the Transmitter (CSIT) Tx_2, i.e. Tx_2 is oblivious of α_{11}, α_{12}, α_{21} and α_{22}.

D. Performances of the primary and secondary protocols

For the rest of this paper, we define the *throughput* as follows.

Definition 1 ([19], [16] or [20]). *The throughput is the average number of information bits correctly received per unit of time.*

For the primary system, the throughput is given by

$$\eta_1(\pi_2) = \lim_{t\to\infty} \frac{1}{t}\mathbb{E}\left(\sum_{n=1}^{t} R_1^n\right) \ bits/cu \tag{3}$$

where t is the time measured in terms of *slots*, R_1^n is a *reward* which is $R_1^n = r_1 \ bits/cu$ if the current packet is successfully decoded after slot n and $R_1^n = 0 \ bits/cu$ if not. $\mathbb{E}\left(\sum_{n=1}^{t} R_1^n\right)$ is the expected number of information bits correctly received per channel uses up to slot t. The notation $\eta_1(\pi_2)$ on the left-hand side of equation (3) is used to enhance the primary throughput dependence on the secondary power and rate allocation noted as

$$\pi_2 = \{(p_2^n, r_2^n)\}_{n\in\mathbb{N}} . \tag{4}$$

The throughput of the secondary protocol can be computed in a similar way. Introducing the secondary one-step reward $R_2^n = r_2^n \ bits/cu$ if the secondary packet is correctly decoded by Rx_2 at the end of slot n and $R_2^n = 0 \ bit/cu$ elsewhere, leads to the following expression for the secondary throughput:

$$\eta_2(\pi_2) = \lim_{t\to\infty} \frac{1}{t}\mathbb{E}\left(\sum_{n=1}^{t} R_2^n\right) . \tag{5}$$

E. Optimization Problem

The secondary user intends to find a joint power and rate allocation $\pi_2 = \{(p_2^n, r_2^n)\}_{n\in\mathbb{N}}$ maximizing the secondary throughput $\eta_2(\pi_2)$ and guaranteeing a target throughput η_{1T} for the primary user. The optimization problem is summarized as follows

$$\begin{gathered} \eta_2^\star = \sup_{\pi_2} \ \eta_2(\pi_2) \\ \text{subject to } \eta_1(\pi_2) \geq \eta_{1T} \end{gathered} \tag{6}$$

III. From System Model to Constrained Markov Decision Process Model

In this section, we present how the HARQ protocol of the PUs can be efficiently represented using Markov chain. This model will then allow us to introduce a Constrained Markov Decision Process (CMDP). We will traduce the constrained optimization problem (6) to an equivalent one in the CMDP framework. We finally give some intrinsic properties of the proposed CMDP that we will use to obtain results on the solution of the (6).

A. Representing HARQ evolution with ACMI

In order to perform its power and rate allocation, Tx_2 is assumed to know only the state of the primary IR-HARQ given by the ACcumulated Mutual Information (ACMI) after k_1 transmissions. The primary ACMI at Rx_1 will be denoted by i_1. We suppose, without loss of generality, that the beginning of those k_1 rounds happens on slot 0. The ACMI can be defined as

$$i_1^{k_1} = \sum_{i=0}^{k_1-1} C\left(\beta_1^i\right), \tag{7}$$

where the function $C(x) = \log_2(1+x)$ is the Shannon capacity of a symmetric Additive White Gaussian Noise (AWGN) channel. The evolution of the HARQ protocol can be fully tracked using the parameters k_1 and $i_1^{k_1}$. Indeed, the decoding failure event after $k_1 < N_1$ transmission is given by the event

$$\mathcal{O}_{k_1} = \left\{ i_1^{k_1} \leq r_1 \right\}.$$

Similarly, the outage event (decoding failure whereas the N_1 codeblocks are sent) is defined as

$$\mathcal{O}_{N_1} = \left\{ i_1^{N_1} \leq r_1 \right\}$$

An illustration of the IR-HARQ evolution of the primary user is given in figure 2. This representation shows how tracking k_1 and i_1 can help determining the current state of the primary HARQ protocol. The HARQ protocol accumulates mutual information until either there is a successful (this event is illustrated in Figure 2) decoding or there is an outage event (this event is not illustrated in Figure 2). The state of the IR-HARQ system before slot n is then fully determined by the couple (k_1^n, i_1^n) where $k_1^n \in \{0, 1 \ldots N_1\}$ represents the number of transmissions performed before slot n. k_1 is set to 0 after a successful decoding and to N_1 after an outage. i_1^n represents the ACMI at Rx_1 before slot n with the convention that after

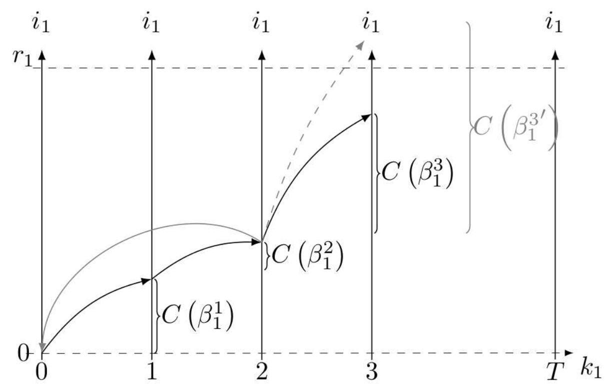

Figure 2. Temporal evolution of the mutual information for the primary IR-HARQ protocol.

a successful decoding or an outage, $i_1^n = 0$. Using the event $\mathcal{O}(n) = \{i_1^n + C\left(\beta_1^n\right) \leq r_1\}$, the evolution from (k_1^n, i_1^n) to (k_1^{n+1}, i_1^{n+1}) can then be given as follows

$$(k_1^{n+1}, i_1^{n+1}) = \begin{cases} (k_1^n + 1, i_1^n + C\left(\beta_1^n\right)) \text{ if } \mathcal{O}(n) \text{ and } k_1^n < N - 1 \\ (N_1, 0) \text{ if } \mathcal{O}(n) \text{ and } k_1^n = N_1 - 1 \\ (1, C\left(\beta_1^n\right)) \text{ if } \mathcal{O}(n) \text{ and } k_1^n = N_1 \\ (0,0) \text{ if } \overline{\mathcal{O}}(n), \end{cases} \tag{8}$$

$\overline{\mathcal{O}}(n)$ describes the event 'not $\mathcal{O}(n)$'.

Because of equation (2), the secondary power and rate allocation (p_2^n, r_2^n) will obviously impact on the evolution of k_1^n and i_1^n. It will then affect the performance of the primary system. Since there is no analytical expression for the primary throughput for all policies (p_2^n, r_2^n), this problem cannot be solved using usual classical tools from optimization theory. However, Constrained Decision Markov Processes (CMDP) seem to be appropriate for solving this problem. Since the space in which the system evolves (space of all (k_1^n, i_1^n)), is neither discrete, nor continuous, we cannot describe its random evolution with neither discrete random variables nor continuous random variables. To circumvent this difficulty we will use the theory of CMDP on Borel Spaces [21] to model and solve the optimization problem (6).

B. Constrained Markov Decision Process

The CMDP definition (see e.g. [12]- [13]) adapted to our problem is a tuple $(\mathbb{S}, \mathbb{A}, \mathbb{W}, Q, R_1, R_2, \eta_{1T})$ where each component is defined as follows.

- The *state space*: $\mathbb{S} = \{0, 1, ..., N\} \times [0, r_1]$ is the set of all possible states of Tx_1. At slot n, $s^n \in \mathbb{S}$ characterized by $s^n = (k_1^n, i_1^n)$ is then observed by Tx_2.
- The *action space*: $\mathbb{A} = [0, P_{2M}] \times [0, R_{2M}]$ is the space of all possible actions available for Tx_2. At slot n Tx_2 will transmit a new packet using a couple power/rate given by $a^n = (p_2^n, r_2^n)$. For the rest of this paper, the set of all couples of states and action is denoted by $\mathbb{K} = \mathbb{S} \times \mathbb{A}$.
- At slot n, the gains α_{11}^n, α_{12}^n, α_{21}^n and α_{22}^n are unknown by Tx_2, they will be considered as disturbances. We consider them as belonging to the *disturbance space* $\mathbb{W} = ([0, +\infty[)^4$.
- The *system function* $g(\cdot)$ is a deterministic function from $\mathbb{S} \times \mathbb{A} \times \mathbb{W}$ to $\mathbb{S}$ which traduces the evolution of the system from state $s_n \in \mathbb{S}$ at slot n to state $s_{n+1} \in \mathbb{S}$ at slot $n+1$ when the action a_n is performed and when the disturbance is w_n. $g(\cdot)$ is defined as follows

$$g(s^n, a^n, w^n) = (k_1^{n+1}, i_1^{n+1}) = s^{n+1}, \tag{9}$$

where k^{n+1} and i_1^{n+1} are given using (8) where w^n is used to compute β_1^n.

- The evolution of the system is statistically represented by the *transition law* denoted by Q and illustrated in figure 3. For a given measurable subset $B \in \mathcal{B}(\mathbb{S})$ and a couple $(s, a) \in \mathbb{K}$, the definition of Q is given by

$$Q(B|s, a) = \mathbb{P}\left(s^{n+1} = g(s^n, a^n, w^n) \in B | s^n = s, a^n = a\right). \tag{10}$$

The expression of $Q(B|s, a)$ is given in Appendix A-A.

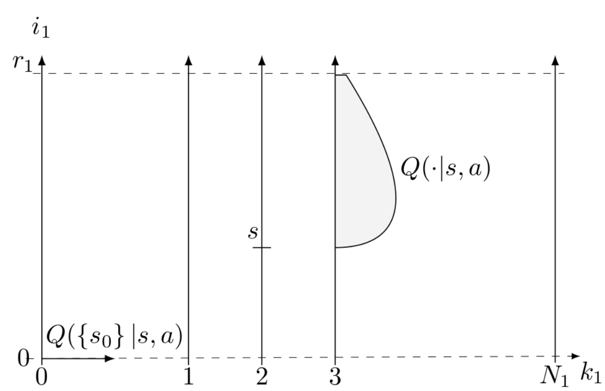

Figure 3. Illustration of the transition law $Q(\cdot|s, a)$, the state $s_0 = (0, 0)$ represents a successful decoding of the packet of the PUs.

- At slot n, the definitions of the primary instantaneous rewards R_1 is given in section II but are rewritten as function of s^n and a^n as follows

$$R_1(s^n, a^n) = r_1 \mathbb{1}_{\{k=0\}}(s^n), \tag{11}$$

where $\mathbb{1}_A(s)$ is the function that is equal to $'1'$ if $s \in A$ and is $'0'$ otherwise.

Similarly $R_2(n)$ can be written as function of $(s^n, a^n) \in \mathbb{K}$ as

$$R_2(s^n, a^n, w^n) = r_2^n \mathbb{1}_{\left\{r_2^n \leq \log_2(1+\beta_2^n)\right\}}(s^n, a^n, w^n). \tag{12}$$

We remark here that R_2 is independent from s^n however, it depends of the action a^n and on the disturbance w^n. As proposed in [22], we will then introduce the average reward as

$$R_2(a^n) = r_2^n \mathbb{P}\left(r_2^n \leq \log_2(1+\beta_2^n)|p_2^n\right). \tag{13}$$

C. Policies

At slot n, we suppose that Tx_2 can store every visited states and every taken actions in a vector called "history" defined as $h_n = (s_0, a_0, s_1, a_1, \ldots, s_{n-1}, a_{n-1}, s_n)$. The space of all possible histories up to time n is recursively defined as $\mathbb{H}_0 = \mathbb{S}$ and $\mathbb{H}_n = \mathbb{K}^n \times \mathbb{S}$.

Suppose now that, at slot n, Tx_2 has stored the history h_n. Accordingly to h_n, Tx_2 randomly chooses an action within the set $\mathbb{A}$. This choice can more formally be written using a the conditional measure $\pi_2^n(\cdot|h_n)$. For any $A \subset \mathbb{A}$, the probability that Tx_2 chooses an action from A $\pi_2^n(A|h_n)$. Since for every time slot n Tx_2 must choose an action, π_2^n verifies that $\pi_2^n(\mathbb{A}|h_n) = 1$. A policy is then defined as a sequence of such $\pi_2 = \{\pi_2^n\}_{n\in\mathbb{N}}$. The set of all the possible policies is denoted by Π.

A policy is said to be *randomized stationary* if there exists a probability measure φ such that for all n, $\pi_2^n(.|h_n) = \varphi(.|s_n)$. The set of all randomized stationary policies is denoted by Π_{RS}.

A policy is said to be *deterministic stationary* if there exists a deterministic function ζ such that for all n, $a^n = \zeta(s^n)$. Using measure notation, it means that for all n, $\pi_2^n(.|h_n) = \delta_{\zeta(s_n)}(.|s_n)$. $\delta_a(B)$ is the Dirac measure, that is the measure that is 1 if $a \in B$ and 0 else. The set of all deterministic stationary policy is denoted by Π_{DS}. Note that we have the following inclusion $\Pi_{DS} \subset \Pi_{RS} \subset \Pi$.

Suppose that the initial state, s_0 is drawn according to some probability distribution ν_0. We will denote by $\mathbb{E}_{\nu_0}^{\pi_2}$ the expectation that is taken over the processes s^n and a^n with respect to the initial distribution ν_0 and the policy π_2.

Into the light of the preceding definitions, one can note that the limits in (3) and (5) are slightly modified as follows

$$\eta_2\left(\pi_2, \nu_0\right) = \liminf_{t\to\infty} \frac{1}{t} \mathbb{E}_{\nu_0}^{\pi_2} \left(\sum_{n=1}^{t} R_2(s_n, a_n) \right) \ bits/cu \tag{14}$$

and

$$\eta_1\left(\pi_2, \nu_0\right) = \liminf_{t\to\infty} \frac{1}{t} \mathbb{E}_{\nu_0}^{\pi_2} \left(\sum_{n=1}^{t} R_1(s_n, a_n) \right) \ bits/cu. \tag{15}$$

Let ν_0 be given, the problem defined by (6) is modified as follows:

$$\begin{aligned} \eta_2^{\star} = \sup_{\pi_2 \in \Pi} \ & \eta_2\left(\pi_2, \nu_0\right) \\ \text{subject to } & \eta_1\left(\pi_2, \nu_0\right) \geq \eta_{1T}. \end{aligned} \tag{16}$$

For the rest of this paper, Ω stands for the set of all admissible policies. Ω is defined as $\Omega = \{\pi \in \Pi | \eta_1\left(\pi_2, \nu_0\right) \geq \eta_{1T}\}$.

IV. Ergodicity results, Consistency and Existence of a solution

In this section, we will prove that every randomized stationary policy induces an ergodic Markov chain on $\mathbb{S}$. This result is then used to prove that every policy in Π is outperformed by a policy in Π_{RS}. This result will allow us to restrain the set of admissible policies Ω to the set of randomized stationary admissible policies $\Omega \cap \Pi_{RS}$. We finally provide a condition on η_{1T} to guarantee that the optimization problem (16) is consistent. In other words, we give a condition on η_{1T} for Ω to be non-empty. We then prove that under the condition of $\Omega \neq \emptyset$ the optimization problem (16) is solvable, that is, there exists $\pi_2 \in \Pi$ such that $\eta_2(\pi_2) = \eta_2^{\star}$.

A. *Ergodicity results*

For every randomized stationary policy φ, the evolution of the state s^n is evolving according to a Markov chain. Indeed, the probability that, at time $n+1$ the system state belongs to a set $B \in \mathcal{B}(\mathbb{S})$ knowing that at time n the system was in state s and that the randomized stationary policy φ is used is given by

$$Q_{\varphi}(B|s) = \int_A Q(B|s,a)\varphi(da|s). \tag{17}$$

In a similar way, for all $t \in \mathbb{N}$, $B \in \mathcal{B}(\mathbb{S})$, $s \in \mathbb{S}$ and $\varphi \in \Pi_{RS}$, we introduce the t step transition probability measure which is the probability that at time $n+t$, $s^{n+t} \in B$ knowing that at time n the system was in state s and φ is applied t times as follows

$$\begin{cases} Q_{\varphi}^{0}\left(B|s\right) = \delta_B(s) \\ Q_{\varphi}^{t}\left(B|s\right) = \int_S Q_{\varphi}^{t-1}\left(B|s'\right) Q_{\varphi}\left(ds'|s\right). \end{cases} \tag{18}$$

We will now show the following theorem

Theorem 1. *Let $\varphi \in \Pi_{RS}$ be given, the Markov chain induced by φ is ergodic. That is, there exists a unique probability measure p_φ verifying*

$$p_\varphi(B) = \int_{\mathbb{S}} Q_\varphi(B|s) p_\varphi(ds),\ \forall B \in \mathcal{B}(\mathbb{S}). \tag{19}$$

Proof: The proof of this theorem, is a direct consequence of the Lemma 3.3 of [21]. This lemma adapted to our case is given here for ease of presentation.

Lemma 1 ([21] Lemma 3.3, p. 57)**.** *If there exists μ, a measure on $\mathbb{S}$ such that*

$$\mu(\mathbb{S}) > 0 \tag{20}$$

$$Q(B|k) \geq \mu(B),\ \forall k \in \mathbb{K},\ \forall B \in \mathcal{B}(\mathbb{S}), \tag{21}$$

then for every $\varphi^\infty \in \Pi_{RS}$ there exists p_φ, a probability measure on $\mathbb{S}$, such that

$$\sup_{s\in\mathbb{S}} \left|\left|Q_\varphi^t(\cdot|s) - p_\varphi(\cdot)\right|\right|_{TV} \underset{t\to\infty}{\to} 0 \tag{22}$$

We further have that p_φ verifies the following property

$$p_\varphi(B) = \int_{\mathbb{S}} Q(B|s,\varphi) p_\varphi(ds),\ \forall B \in \mathcal{B}(\mathbb{S}).$$

It then remains to prove that there exists a probability measure μ satisfying equations (20) and (21). To do so, remark that the state $s_0 = (0,0)$, is accessible from every other states. s_0 physically represents a successful decoding of Rx_1. The fact that s_0 is accessible form every other states means that from every state s, there is a non-zero probability of a successful decoding at Rx_1. Consider a state $s \in \mathbb{S}$ and an action $a \in \mathbb{A}$, we have the following

$$\begin{aligned} Q(\{s_0\}|s,a) &= \mathbb{P}\left(i_1 + C(\frac{p_1\alpha_{11}}{1+p_2\alpha_{21}}) \geq r_1 | i_1, p_2\right) \\ &\geq \mathbb{P}\left(C(\frac{p_1\alpha_{11}}{1+P_{2M}\alpha_{21}}) \geq r_1\right) > 0. \end{aligned} \tag{23}$$

Let a set $B \in \mathcal{B}(\mathbb{S})$ be given, for every $k \in \mathbb{K}$, we have

$$\begin{aligned} Q(B|k) &= Q(B \cap \{s_0\}|k) + Q(B \cap \{s_0\}|k) \\ &\geq Q(B \cap \{s_0\}|k) \\ &= \mathbb{P}\left(C(\frac{p_1\alpha_{11}}{1+p_{2M}\alpha_{21}} \geq r_1)\right)\mathbb{1}_B(s_0) \end{aligned} \tag{24}$$

For all $B \in \mathcal{B}(\mathbb{S})$ consider $\mu(B) = \mathbb{P}\left(C(\frac{p_1\alpha_{11}}{1+p_{2M}\alpha_{21}} \geq r_1)\right)\mathbb{1}_B(s_0)$. We finally have that for every $k \in \mathbb{K}$ and every $B \in \mathcal{B}(\mathbb{S})$, $Q(B|k) \geq \mu(B)$ furthermore $\mu(\mathbb{S}) = \mathbb{P}\left(C(\frac{p_1\alpha_{11}}{1+P_{2M}\alpha_{21}}) \geq r_1\right) > 0$, where the last inequality comes from the fact that P_{2M} is bounded. Equations (20) and (21) are both verified which proves that every randomized stationary policies induces an ergodic Markov chain on $\mathbb{S}$. ■

These ergodicity results also leads to the fact that for every randomized stationary policy, the primary and secondary throughputs can be written as follows

$$\eta_i(\nu_0, \varphi) = \int_S R_i(s,\varphi)p_\varphi(ds), \ i \in \{1,2\}. \tag{25}$$

Remark 1. *For every $\varphi \in \Pi_{RS}$ and for $i = 1$ and $i = 2$, the right hand side of equation* (25)*, does not depend on the initial distribution ν_0. In the sequel, for every φ we will then write*

$$\eta_i(\varphi) = \eta_i(\nu_0, \varphi), \tag{26}$$

$$= \int_S R_i(s,\varphi)p_\varphi(ds). \tag{27}$$

B. Domination of the randomized stationary policies

The ergodicity property developed in the previous section, especially the result given by equation (25) have been used in [15] in order to show the following Lemma.

Lemma 2 ([15], lemma 3.5 p. 448)**.** *If Ω is non-empty, then the following holds $a)$ For every initial distribution ν_0 and for every $\pi \in \Omega$, there exists $\varphi \in \Pi_{RS}$ such that*

- *$\varphi \in \Omega$ and*
- $\eta_2(\varphi) \geq \eta_2(\pi, \nu_0)$

Proof: The proof of Theorem 3, is similar to the one proposed in [23] for the general case of CMDP. We then emphasize the main differences. The proof directly build a randomized stationary policy. Define the occupation measure as follows

$$m_n(\Gamma) = \frac{1}{n}\sum_{t=0}^{n-1} \mathbb{P}_{\nu_0}^{\pi_2}\left((s^n, a^n) \in \Gamma\right),\ \Gamma \in \mathcal{B}(\mathbb{S} \times \mathbb{A}),$$

and express η_1 and η_2 as

$$\eta_i(\pi, \nu_0) = \liminf_{n\to\infty} \int_{\mathbb{S}\times\mathbb{A}} R_i(s,a)m_n(ds,da).$$

Using the compactness of the space $\mathbb{S}$ in the theorem of Prohorov given in Appendix B-A gives that there exists a measure m and a subsequence n_j such that $m_{n_j} \to m$. Using the continuity of R_1 and R_2 gives the following result

$$\eta_i(\pi, \nu_0) = \lim_{j\to\infty} \int_{\mathbb{S}\times\mathbb{A}} R_i(s,a)m_{n_j}(ds,da) = \int R_i dm.$$

They then show that m can be disintegrated in $m = \varphi p_\varphi$ (see appendix B-B for the disintegration result) where p_φ verifies equation (19). This justifies the fact that m define the required randomized stationary policy. ■

Lemma 2 means that for every policy $\pi \in \Omega$, there exists a randomized stationary policy $\varphi \in \Pi_{RS} \cap \Omega$ that performs better. We then say that Π_{RS} dominates Π. The optimization problem (16) can then be rewritten in Π_{RS} as follows

$$\eta_2^\star = \sup_{\varphi\in\in\Pi_{RS}\cap\Omega} \eta_2(\varphi) \tag{28}$$

For the rest of this paper, we denote the set of feasible randomized stationary policies by $\Omega_{RS} = \Pi_{RS} \cap \Omega$.

C. Consistency of the optimization problem (28)

Let η_{10} be the throughput of the primary user when the secondary user uses the allocation ζ_0 defined for all $s \in \mathbb{S}$ as $\zeta_0(s) = (0,0)$. We giver hereafter a condition of consistency for (28).

Theorem 2 (Consistency of the optimization problem (28))**.**
The constrained problem (28) *is consistent if and only if* $\eta_{1T} \leq \eta_{10}$. *By definition, the constrained problem* (28) *is said to be consistent if and only if the set of all* Ω_{RS} *is non-empty.*

We will now prove a lemma that will be used in order in the proof of 2.

Lemma 3. *Let $\varphi_1 \in \Pi_{RS}$ and $\varphi_2 \in \Pi_{RS}$ be two given policies. Suppose that for every $s \in \mathbb{S}$, there exists $K_1(s) \subset \mathbb{A}$ and $K_2(s) \subset \mathbb{A}$ such that*

$$\varphi_1(K_1(s)|s) = 1 \tag{29}$$

$$\varphi_2(K_2(s)|s) = 1 \tag{30}$$

$$\forall a_1 = (p_2^1, r_2^1) \in K_1(s),\ a_2 = (p_2^2, r_2^2) \in K_2(s),\ a_1 \succeq a_2, \tag{31}$$

where we write $a_1 \succeq a_2$ if either '$p_2^1 > p_2^2$' or '$p_2^1 = p_2^2$ and $r_2^1 \geq r_2^2$'. Under these conditions, we have the following inequality $\eta_1(\varphi_1) \leq \eta_1(\varphi_2)$.

Proof: Let φ_1 and φ_2, verifying the hypotheses of proposition 3. Since both policies belong to Π_{RS}, their primary throughput can be written as follow $\eta_1(\varphi_1) = \int_S R_1(s, \varphi_1) p_{\varphi_1}(ds)$ and $\eta_1(\varphi_2) = \int_S R_1(s, \varphi_2) p_{\varphi_2}(ds)$, which, using the definition of $R_1(s, a)$ given in equation (11) can be written as $\eta_1(\varphi_1) = R_1 p_{\varphi_1}(\{s_0\})$ and $\eta_1(\varphi_2) = R_1 p_{\varphi_2}(\{s_0\})$. Also, for $i \in \{1, 2\}$, $p_{\varphi_i}(\{s_0\})$ have to verify equation (19), which in this case can be written as follows

$$p_{\varphi_i}(\{s_0\}) = \int_{\mathbb{S}} Q_{\varphi_i}(\{s_0\}|s) p_{\varphi_i}(ds),$$

For every $s \in \mathbb{S}$, $Q_{\varphi_i}(\{s_0\}|s)$ is evaluated as follows

$$Q_{\varphi_i}(\{s_0\}|s) = \int_{\mathbb{A}} \mathbb{P}\left(i_1 + C\left(\frac{p_1\alpha_{11}}{1 + p_2\alpha_{21}}\right) \geq R_1 | s, a\right) \varphi_i(da|s),$$

Due to equation (31), we have that for all $s \in \mathbb{S}$, $a_1 = (p_2^1, r_2^1) \in K_1(s)$ and $a_2 = (p_2^2, r_2^2) \in K_2(s)$, we have that $p_2^1 \geq p_2^2$. This obviously implies the following inequality

$$\mathbb{P}\left(i_1 + C\left(\frac{p_1\alpha_{11}}{1 + p_2^1\alpha_{21}}\right) \geq R_1 | s, a_1\right) \leq \mathbb{P}\left(i_1 + C\left(\frac{p_1\alpha_{11}}{1 + p_2^2\alpha_{21}}\right) \geq R_1 | s, a_2\right).$$

We finally have that for all $s \in \mathbb{S}$, $Q_{\varphi_1}(\{s_0\}|s) \leq Q_{\varphi_2}(\{s_0\}|s)$ which proves that $\eta_1(\varphi_1) \leq \eta_1(\varphi_2)$. ■

Proof of theorem 2: The proof of the direct part, that is "$\eta_{1T} \leq \eta_{10} \Rightarrow \Omega_{RS} \neq \emptyset$", is obvious. Indeed if $\eta_{1T} \leq \eta_{10}$ implies that $\zeta_0 \in \Omega_{RS}$ which in turn implies that $\Omega_{RS} \neq \emptyset$.

Let show now the converse part, that is, "$\Omega_{RS} \neq \emptyset \Rightarrow \eta_{1T} \leq \eta_{10}$". Suppose that $\Omega_{RS} \neq \emptyset$. This implies that there exists $\varphi \in \Omega_{RS}$. Due to the definition of a randomized stationary policy, φ must verify that for every $s \in \mathbb{S}$, $\varphi(\mathbb{A}|s) = 1$. Applying Lemma (3) with $\varphi_1 = \varphi$, $\varphi_2 = \zeta_0$

and for every $s \in \mathbb{S}$ $K_1(s) = \mathbb{A}$, $K_2(s) = (0,0)$ leads to the conclusion that $\eta_1(\varphi) \leq \eta_1(\zeta_0)$. Considering now that $\eta_{10} = \eta_1(\zeta_0)$ and that $\varphi \in \Omega_{RS}$, we have

$$\eta_{1T} \leq \eta_1(\varphi) \leq \eta_{10}$$

which conclude this proof. ■

Theorem 2 stands that the set of admissible policies is non-empty if and only if the PUs throughput constraint is below the throughput of the primary system in absence of SUs. This conclusion means that no policy in Π can improve the throughput of the PUs which is logical for our model.

D. Solvability of the optimization problem

In this section, we show that there exists a solution to the optimization problem (28). That is, we want to show the following theorem.

Theorem 3 (Solvability)**.** *If There exists $\varphi \in \Pi_{RS}$ such that, $\eta_2(\varphi) = \eta_2^\star$ and $\eta_1(\varphi) \geq \eta_{1T}$.*

Proof: The proof of this theorem is again similar to the one given in [15] considering the same modifications as the one done in the proof of Lemma 2. ■

In this section, we have shown that the optimization problem (16) can be reduced from Π to Π_{RS}, without loss of optimality. We have then given a condition on the primary throughput constraint that guarantees the consistency of the optimization problem (16). Under this consistency condition, we have finally shown that there exists an optimal randomized strategy. In the next section we will show how a linear programming approach can be used to give an algorithm that compute an optimal policy.

V. The linear programming approach

In this section, we show that the optimization problem (16) can be viewed as a linear programming in infinite dimensional space. We then give the dual formulation of the linear programming, and based on this dual formulation, we propose an algorithm to build an optimal policy.

A. Linear Programming Formulation

In section IV, we have shown that a probability measure m on $\mathbb{S} \times \mathbb{A}$ define a randomized stationary policy φ if $m = \varphi\hat{m}$ and $\hat{m}$ verifies equation (19). This can be traduced as follows: m is a probability measure if and only if

$$\int_{\mathbb{S}\times\mathbb{A}} m(d(s,a)) = 1, \tag{32}$$

and m define some φ if and only if

$$L_0 m(B) = \hat{m}(B) - \int_{\mathbb{S}\times\mathbb{A}} Q(B|s,a)m(d(s,a)) = 0,\ \forall B \in \mathcal{B}(\mathbb{S}). \tag{33}$$

We have also shown that the primary and secondary throughputs can be written as follows

$$\eta_i(\varphi) = \int_{\mathbb{S}\times\mathbb{A}} R_i(s,a)m(d(s,a)),\ i \in \{1,2\}. \tag{34}$$

Consider the following two dual pairs of vector spaces $(\mathcal{M}(\mathbb{K}), \mathcal{F}(\mathbb{K}))$ and $(\mathcal{M}(\mathbb{S}), \mathcal{F}(\mathbb{S}))$. $\mathcal{M}(\mathbb{K})$ (resp. $\mathcal{M}(\mathbb{S})$) is the space of signed measures on the space $\mathbb{K}$ (resp. on the space $\mathbb{S}$). $\mathcal{F}(\mathbb{K})$ (resp. $\mathcal{F}(\mathbb{S})$) is the space of bounded measurable functions on $\mathbb{K}$ (resp. on the space $\mathbb{S}$). Let $\langle\cdot,\cdot\rangle_{\mathbb{K}}$ be a bilinear form for the dual pair $(\mathcal{M}(\mathbb{K}), \mathcal{F}(\mathbb{K}))$ defined as follows

$$\langle m, v\rangle_{\mathbb{K}} = \int_{\mathbb{K}} v(s,a)m(d(s,a)). \tag{35}$$

In a similar way, we introduce $\langle\cdot,\cdot\rangle_{\mathbb{S}}$ for the dual pair $(\mathcal{M}(\mathbb{S}), \mathcal{F}(\mathbb{S}))$. The optimization problem (28) can then be rewritten as

$$\begin{aligned}
\eta_2^\star = &\sup_{m\in\mathcal{M}(\mathbb{K})^+} \langle m, R_2\rangle_{\mathbb{K}}, \\
s.t. &L_0 m(B) = 0,\ \forall B \in \mathcal{B}(\mathbb{S}), \\
&\langle m, 1\rangle_{\mathbb{K}} = 1, \\
&\langle m, R_1\rangle_{\mathbb{K}} \geq \eta_{1T},
\end{aligned} \tag{36}$$

where $\mathcal{M}(\mathbb{K})^+$ is the cone of positive measures on $\mathbb{K}$. Since L_0 is a linear map from $\mathcal{M}(\mathbb{K})$ to $\mathcal{M}(\mathbb{S})$, its adjoint L_0^* is the map such that for every $v \in \mathcal{F}(\mathbb{S})$

$$\langle L_0 m, v\rangle_{\mathbb{S}} = \langle m, L_0^* v\rangle_{\mathbb{K}}, \tag{37}$$

and is given by the following equation

$$(L_0^* v)(s,a) = v(s) - \int_{\mathbb{S}} v(y)Q(dy|s,a),\ \forall (s,a) \in \mathbb{K},\ v \in \mathcal{F}(\mathbb{S}). \tag{38}$$

Since $v \in v \in \mathcal{F}(\mathbb{S})$ and Q is strongly continuous, we have that $L_0^* v \in v \in \mathcal{F}(\mathbb{K})$ and we have that L_0^* is a linear map from $\mathcal{F}(\mathbb{S})$ to $\mathcal{F}(\mathbb{K})$ which guarantees that L_0 is continuous with respect to the weak topology (cf. [15]) and shows that (36) is a linear program. The constraint $\langle m, R_1 \rangle_{\mathbb{K}} \geq \eta_{1T}$ can classically be reduced to an equality constraint using a slack variable $\alpha \in \mathbb{R}^+$ such that $\langle m, R_1 \rangle_{\mathbb{K}} + \alpha = \eta_{1T}$. We then extend $\langle \cdot, \cdot \rangle_{\mathbb{K}}$ to $(\mathcal{M}(\mathbb{K}) \times \mathbb{R}, \mathcal{F}(\mathbb{K}) \times \mathbb{R})$ as follows

$$\langle (m, x), (v, y) \rangle = \langle m, v \rangle_{\mathbb{K}} + xy. \tag{39}$$

We then obtain the following equality constrained linear programming

$$\begin{aligned}
\eta_2^\star = &\sup \langle (m, \alpha), (R_2, 0) \rangle = \int R_2 dm + \alpha.0, \\
s.t.\ &L_0 m(B) = 0,\ \forall B \in \mathcal{B}(\mathbb{S}) \\
&\langle m, 1 \rangle_{\mathbb{K}} = 1 \\
&\langle m, R_1 \rangle_{\mathbb{K}} + \alpha = \eta_{1T}, \\
&m \in \mathcal{M}_+(\mathbb{K}),\ \alpha \in \mathbb{R}^+
\end{aligned} \tag{40}$$

Since the linear programming (40) is equivalent to the initial optimization problem (28), it is consistent and solvable. The dual linear programming of (40) can be written as follows

$$\begin{aligned}
\eta^\star = &\inf \eta - \lambda \eta_{1T}, \\
s.t.\ &\eta + u(s) \geq R_2(s, a) + \lambda R_1(s, a) + \int_{\mathbb{S}} u(y) Q(dy|s, a) \\
&u \in \mathcal{F}(\mathbb{S}),\ \eta \in \mathbb{R},\ \lambda \in \mathbb{R}^+.
\end{aligned} \tag{41}$$

We can already note that the dual programming (41) is consistent since $(u = 0, \eta = r_{2M}, \lambda = 0)$ is admissible.

B. *Absence of duality gap and strong duality*

The two linear programmings given in (40) and (41) are both consistent and (40) is solvable. We then have that there exists a randomized stationary policy $\varphi^\star \in \Omega$ such that $\eta_2^\star = \eta_2(\varphi^\star)$. Also since the linear programming (41) is the dual programming of (40), the consistencies imply weak duality, that is $\eta_2^\star \leq \eta^\star$.. For the rest of this section, we will then suppose that $\eta_{1T} < \eta_{10}$. This condition implies the consistency of the primal optimization problem (40).

We will prove that there is no duality gap between the primal linear programming (40) and its dual (41), that is $\eta_2^\star = \eta^\star$. We further prove that (41) is solvable. That is, there exists a triplet (η, λ, u) admissible for the linear programming (41) such that $\eta = \eta^\star$.

Theorem 4 (Absence of duality gap)**.** *There is no duality gap, that is* $\eta^\star = \eta_2^\star$.

Proof: For ease of notation, we introduce the linear map $L(m, \alpha) = (L_0 m, \langle m, R_1\rangle_{\mathbb{K}} + \alpha, \langle m, 1\rangle_{\mathbb{K}})$. The proof is realized by showing that the set

$$H = \left\{ \left(L(m,\alpha), \langle (m,\alpha), R_2, 0\rangle + r \right), m \in \mathcal{M}(\mathbb{K})_+, \alpha \in \mathbb{R}^+, r \in \mathbb{R}^+ \right\}$$

is closed (cf. [24]). ■

We now show that there exists $(u^\star, \eta^\star, \lambda^\star)$ solution to the linear programming (41).

Theorem 5 (Strong duality)**.** *The optimization problem* (41) *is solvable. There exists a triplet* $(u^\star, \eta^\star, \lambda^\star)$ *feasible for* (41).

Proof: First, remark that we have

$$\eta^\star = \inf_{\lambda \in \mathbb{R}^+} \eta_\lambda^\star - \lambda \eta_{1T}, \tag{42}$$

where for a given $\lambda > 0$, $\eta_\lambda^\star$ is given by the following equation

$$\begin{aligned} \eta_\lambda^\star = & \inf_{u \in \mathcal{F}(\mathbb{S}),\ \eta \in \mathbb{R}} \eta \\ & s.t.\ \eta + u(s) \geq R_2(s,a) + \lambda R_1(s,a) + \int_{\mathbb{S}} u(y) Q(dy|s,a). \end{aligned} \tag{43}$$

We will then proceed in two steps. (a) we will show that for every $\lambda \in \mathbb{R}^+$, there exists (η, u) solution of (43), (b) we will show that there exists λ solution of (42).

As it has been shown in [15], solving (43) is equivalent to solving an unconstrained MDP with one-step reward function $R_2(s,a) + \lambda R_1(s,a)$ keeping unchanged $\mathbb{S}$, $\mathbb{A}$ and Q. Indeed, consider the following unconstrained MDP

$$\eta_\lambda' = \sup_{\varphi \in \Pi_{RS}} \eta_\lambda(\varphi), \tag{44}$$

where $\eta_\lambda(\varphi)$ is given by the following expression $\eta_\lambda(\varphi) = \eta_2(\varphi) + \lambda \eta_1(\varphi)$. Using the same argument as before, we remark that the optimization problem (43) is the dual of the optimization problem (44). Since both problems are consistent, the weak duality gives the following inequality

$\eta'_\lambda \leq \eta^\star_\lambda$. Using the ergodicity condition given in Lemma 1, it is shown in [21] that there exists a constant η and a bounded function u such that

$$\eta + u(s) = \sup_{a\in\mathbb{A}} \left\{ R_2(s,a) + \lambda R_1(s,a) + \int_{\mathbb{S}} u(y)Q(dy|s,a) \right\} \;\; \forall s \in \mathbb{S}. \tag{45}$$

They have also show that if the couple $(\eta^\star_\lambda, u_\lambda)$ satisfies equation (45) then $\eta = \eta'_\lambda$. Remarking that the couple (η, u) defined by equation (45) is admissible for the optimization problem (43) leads to the inequality $\eta'_\lambda \geq \eta^\star_\lambda$. This justifies that $(\eta^\star_\lambda, u_\lambda) = (\eta, u)$.

We will now show that there exists $\lambda \geq 0$ optimal for (42). For a given $\lambda \geq 0$, let ζ_λ be such that for every $s \in \mathbb{S}$, $\zeta_\lambda(s)$ is an argument of the maximum of problem (43). The existence of such ζ_λ is guaranteed by the compactness of $\mathbb{A}$, by the continuity of R_1 and R_2 and by the strong continuity of Q (see Appendix A-C). We denote by $\eta_1^\lambda = \eta_1(\zeta_\lambda)$ and by $\eta_2^\lambda = \eta_2(\zeta_\lambda)$ the primary and secondary throughputs. With trivial adaptations of the results of [13], we show that η_1^λ and $\eta^\star_\lambda$ are increasing functions of λ and that η_2^λ is a decreasing function of λ. We can also show that the function $\eta^\star_\lambda$ is absolutely continuous. By absolute continuity of $\eta^\star_\lambda$, its right derivative must coincide with the ordinary derivative, that is

$$\frac{d\eta^\star_\lambda}{d\lambda} = \left(\frac{d\eta^\star_\lambda}{d\lambda}\right)^+ = \eta_1^\lambda.$$

Let the function $w(\lambda)$ be defined as follows $w(\lambda) = \eta^\star_\lambda - \lambda\eta_{1T}$. The expression of $w(\lambda)$ is not known thus it cannot be used directly to find the optimal value for λ. Also, $w(\lambda)$ is not differentiable for every λ. However $w(\lambda)$ is differentiable almost everywhere and we have

$$\frac{dw(\lambda)}{d\lambda} = \eta_1^\lambda - \eta_{1T}.$$

Since we have shown that η_1^λ is an increasing (non-necessary continuous) function, $w(\lambda)$ is a convex function and then possesses a unique minimum. If we have that $\eta_{1T} < \eta_{10}$, we have that there exists $\lambda^\star$ such that $\eta_1^\lambda - \eta_{1T} \leq 0$ if $\lambda \leq \lambda^\star$ and $\eta_1^\lambda - \eta_{1T} \geq 0$ if $\lambda \geq \lambda^\star$. This proves that $\lambda^\star$ is such that $\eta^\star = \eta_{\lambda^\star} - \lambda^\star\eta_{1T}$. ∎

That far, we have proven that there is no duality gap between the linear programming (40) and its dual given in equation (41). We have also shown that there exists a triplet (η, λ, u) which feasible for the optimization problem (41). This triplet can be found with the structure $(\eta^\star_{\lambda^\star}, \lambda^\star, u_{\lambda^\star})$.

C. A dual based algorithm for finding an optimal policy

In this section, we propose an algorithm based on the dual programming (41) for finding an optimal solution of the linear programming (40). This condition implies the consistency of the problem (16). We have shown that there exists an optimal policy for the optimization problem (16). Let $\varphi^\star$ be an optimal policy for the optimization problem (16).

For every λ, we determine $(\eta_\lambda, u_\lambda)$ such that $(\eta_\lambda, \lambda, u_\lambda)$ is feasible for the dual programming (41) using the Relative Value Iteration (RVI) algorithm [14]. The RVI algorithm is given as follows.

Algorithm 1 Relative Value Iteration (VI) Algorithm

1: $u^{(0)}(s) \leftarrow 0,\ \forall s \in \mathbb{S}$
2: **for** $k = 1$ to ∞ **do**
3: $\quad \forall s \in \mathbb{S},\ u^{(k)}(s) \leftarrow \max_{a\in\mathbb{A}}\ R_2(s,a) + \lambda R_1(s,a) + \int_{\mathbb{S}} u^{(k-1)}(s')Q(ds'|s,a)$
4: $\quad \forall s \in \mathbb{S},\ u^{(k)}(s) \leftarrow u^{(k)}(s) - u^{(k)}(s_0)$
5: **end for**
6: $\forall s \in \mathbb{S}\ \zeta_\lambda(s) \in \arg\max_{a\in\mathbb{A}}\ R_2(s,a) + \lambda R_1(s,a) + \int_{\mathbb{S}} u_\lambda(s')Q(ds'|s,a)$
7: $u_\lambda = u^{(\infty)}$
8: $\eta_\lambda = \max_{a\in\mathbb{A}}\ R_2(s,a) + \lambda R_1(s,a) + \int_{\mathbb{S}} u_\lambda(s')Q(ds'|s_0,a)$

[21] shows that under the ergodicity results of lemma 1, this algorithm converges.

The compactness of $\mathbb{A}$ and the continuity of the function $U_\lambda(s,a) = R_2(s,a) + \lambda R_1(s,a) + \int_{\mathbb{S}} u_\lambda(s')Q(ds'|s,a)$, ensure that for every $s \in \mathbb{S}$, $K_\lambda(s) = \arg\max_{a\in\mathbb{A}} U_\lambda(s,a)$ is a non-empty compact set. We can then build ζ_λ for every $\lambda \geq 0$. This allows to compute η_1^λ for every λ. We then find $\lambda^\star$ with a standard dichotomy algorithm.

We will now use the following theorem of [13] to find an optimal policy for the optimization problem (16).

Theorem 6. *Let $\varphi \in \Pi_{RS}$, if $\eta_{\lambda^\star}(\varphi) = \eta^\star_{\lambda^\star}$ and $\eta_1(\varphi) = \eta_{1T}$ then φ is optimal for the optimization problem* (16)*.*

Obviously in our case, if $\eta_1(\zeta_{\lambda^\star}) = \eta_{1T}$ then $\zeta_{\lambda^\star}$ is optimal. Suppose now that $\eta_1(\zeta_{\lambda^\star}) \neq \eta_{1T}$. We will construct a policy verifying the conditions of Theorem 6. Since the (28) is solvable,

there exists an optimal policy denoted by $\varphi^\star$. Furthermore we have that $\varphi^\star$ verifies that for every $s \in \mathbb{S}$

$$\begin{aligned}\eta_{\lambda^\star} + u_{\lambda^\star}(s) &= R_2(s,\varphi^\star) + \lambda^\star R_1(s,\varphi^\star) + \int_{\mathbb{S}} u_{\lambda^\star}(s')Q(ds'|s,\varphi^\star),\\ &= \max_{a\in\mathbb{A}} R_2(s,a) + \lambda^\star R_1(s,a) + \int_{\mathbb{S}} u_{\lambda^\star}(s')Q(ds'|s,a)\end{aligned}$$

This implies that for every $s \in \mathbb{S}$, $\varphi(K_{\lambda^\star}(s)|s) = 1$. It means that $s \in \mathbb{S}$, the policy $\varphi^\star$ takes its actions only among $K_{\lambda^\star}(s)$. Consider $\mathbb{A}$ with the order defined by $\succeq$ in Lemma 3. Since $K_{\lambda^\star}(s)$ is compact, it possesses a maximal element for the relation $\succeq$, that is, there exist $\overline{a}(s)$ such that for all $a \in K_{\lambda^\star}(s)$, $\overline{a}(s) \succeq a$. Similarly, we can define $\underline{a}(s)$ as the minimal element of $K_{\lambda^\star}(s)$. For all $s \in \mathbb{S}$ take $\zeta^+(s) = \underline{a}(s)$ and $\zeta^-(s) = \overline{a}(s)$. By construction of ζ^+ and ζ^- and due to the lemma 3 we have that $\eta_1(\zeta^-) \leq \eta_1(\varphi) \leq \eta_1(\zeta^+)$.

Also, we can easily verify that $\eta^\star_{\lambda^\star} = \eta_{\lambda^\star}(\zeta^+) = \eta_{\lambda^\star}(\zeta^-)$. Using now the same continuity arguments as in [13], the function $\beta \mapsto \eta_1(\beta\zeta^+ + (1-\beta)\zeta^-)$ is a continuous function of β. Since for $\beta = 0$ we obtain $\eta_1(\zeta^-) \leq \eta_{1T}$ and for $\beta = 1$ we obtain $\eta_1(\zeta^+) \geq \eta_{1T}$, there exists $\beta^\star$ such that $\eta_1(\beta^\star\zeta^+ + (1-\beta^\star)\zeta^-) = \eta_{1T}$. The policy $\beta^\star\zeta^+ + (1-\beta^\star)\zeta^-$ verifies the assumptions of Theorem 6, it is then optimal for 28.

In this section, we have shown that it is possible to find an optimal policy as a mix between two deterministic stationary policies. Those two policies are computed with a dynamic programming called Relative Value Iteration.

VI. Simulation Results and discussion

A. Simulation Results

In this section, some simulation results are proposed. We consider that the PUs are using an IR-HARQ protocol with a maximum of 4 retransmission so that $N_1 = 5$. The equivalent rate of the first block is $r_1 = 5$ *bpcu*. Tx_1 is communicating with a normalized power $p_1 = 3.16$ (without unit) which corresponds to $5dB$. This leads to the state space $\mathbb{S} = \{0, 1 \cdots 5\} \times [0, 5]$.

The SUs can use normalized powers within the set $[0, 10]$ (without unit). The set of available rates is $[0, 4]$ (bpcu). This leads to an action space $\mathbb{A} = [0, 10] \times [0, 4]$.

The parameters of the exponential fading are taken as follows: $\lambda_{11} = \lambda_{21} = \lambda_{12} = \lambda_{22} = 1$. This completely define $\mathbb{W}$.

In general, no closed-form expression can be given for the equations of the RVI algorithm (Algorithm 1). We will then approximate the continuous solution by quantifying the state space $\mathbb{S}$ with 100 values linearly spaced,

$$\hat{\mathbb{S}} = \{0, 1, \cdots, 5\} \times \{0, \epsilon_i, 2\epsilon_i, \cdots, 99\epsilon_i\},$$

where ϵ_i is chosen such that $\left\lfloor \frac{5}{\epsilon_i} \right\rfloor = 99$. Similarly, we quantify the space p_2 as follows $\{0, \epsilon_p, 2\epsilon_p, \cdots, 99\epsilon_p\}$.

One can remark that the optimizations of the Algorithm 1 can be done in a first time on r_2 and in a second time on p_2. The action space $\mathbb{A}$ is then quantified using the following set

$$\hat{\mathbb{A}} = \{(0,0), (\epsilon_p, r_2^\star(\epsilon_p)), (2\epsilon_p, r_2^\star(2\epsilon_p)), \cdots, (99\epsilon_p, r_2^\star(99\epsilon_p))\},$$

where $r_2^\star(p_2) = \max_{r_2} R_2(p_2, r_2)$. The function $r_2^\star(p_2)$ is given in Figure 4 for $p_2 \in [0, 10]$.

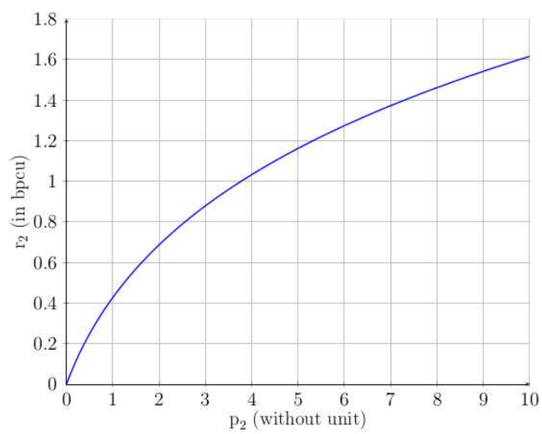

Figure 4. $r_2^\star$ versus p_2

We then use $\hat{\mathbb{S}}$ and $\hat{\mathbb{A}}$ to replace the integral by a sum in Algorithm 1. We give the result obtained for $\lambda = 0.4755$ in Figure 5. To summarize the results obtain for different values of η_{1T}, we introduce the *throughput region*. The throughput region is the set of every couples (η_1, η_2) achievable while using an allocation. In Figure 6, we give the throughput region corresponding to the allocations computed using Algorithm 1.

B. Discussion

• It is important to remark that quantifying $\mathbb{S}$ and $\mathbb{A}$ can lead to bad approximations of the continuous solution. However, it can be shown that due to the Lipschitz continuity of the reward

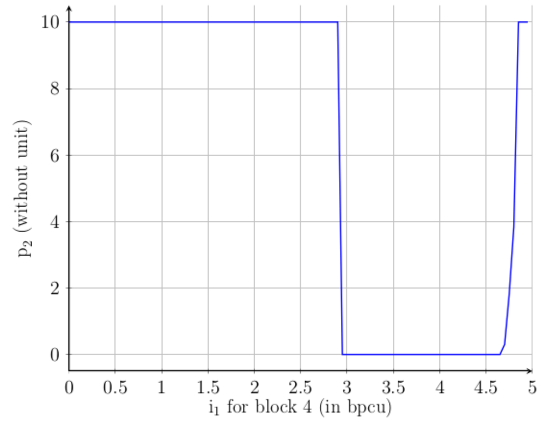

Figure 5. Example of a power allocation for the forth block of the IR-HARQ protocol. This result has been obtained for $\lambda = 0.4755$.

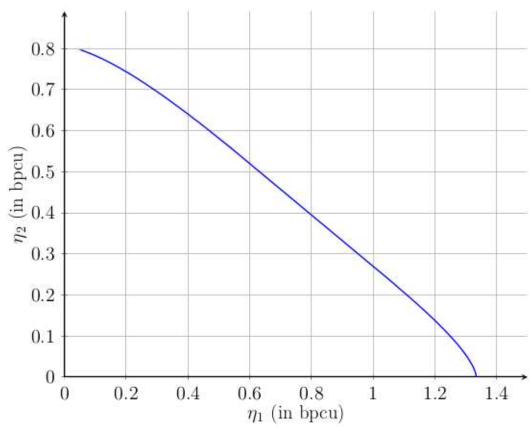

Figure 6. Throughput region of the proposed power and rate allocation.

functions R_1 and R_2, due to the Lipschitz continuity of the transition kernel Q and due to the ergodicity results, finer quantizations of $\mathbb{S}$ and $\mathbb{A}$ lead to better approximations η_λ and ζ_λ. Since the strategies computed with the quantified versions of $\mathbb{S}$ and $\mathbb{A}$ are admissible for the original problem, the corresponding throughput region is a inner bound of the "true" throughput region. Effects of the quantification have been studied in [21] or [25].

- One can also note that Tx_2 must know $s = (k_1, i_1)$ in order to compute the proposed power and rate allocation. The problem we consider is, in that sense, completely observable (CO-MDP). However, note that k can be tracked easily by "counting" the feedbacks of the PUs. The knowledge of i_1 can be more difficult to acquire. One can then use a partially observable

model (PO-MDP). PO-MDP are presented in [22] or [21]. It is well known that PO-MDP can be difficult to solve however some approximations are proposed in [26].

VII. Conclusion

In this paper, we proposed to realize a secondary rate and power allocation when the SUs are aware of the parameters describing the PUs IR-HARQ protocol. We have shown that the ACMI model used to study the IR-HARQ protocol imposes a CMDP formulation for the allocation problem. Based on this CMDP formulation, we have also given conditions for the allocation problem to be consistent. We have shown that under the condition of consistency, an optimal allocation exists among the class of the randomized stationary policies. Taking into account the structure of the optimal policy, we have shown that this former is a solution of an infinite dimensional linear programming. We then solved the linear programming via its dual which is solved using the RVI algorithm. We finally propose and comment some simulation results.

The quantifications of the state and action spaces imply an increasing of the complexity of the RVI. However, lower complexity algorithms can be used to learn suboptimal policies. Future work will then be dedicated to the study of approximations leading to these learning algorithms.

Appendix A
Properties of the CMDP

A. *Expression of* $Q(B|s,a),\ B \in \mathcal{B}(\mathbb{S}),\ s \in \mathbb{S}$ *and* $a \in \mathbb{A}$

Fix $n \geq 0$, $(s,a) \in \mathbb{K}$ and $B \in \mathcal{B}(\mathbb{S})$, the definition of $Q(B|s,a)$ is given by equation (10) that we rewrite for sake of clarity

$$Q(B|s,a) = \mathbb{P}\left(s_{n+1} \in B | s^n = s, a^n = a\right),$$

where $s^{n+1} = (k_1^{n+1}, i_1^{n+1})$, $s^n = (k_1^n, i_1^n)$ and $a^n = (p_2^n, r_2^n)$. Keeping these notations and adding the disturbance $w^n = (\alpha_{11}^n, \alpha_{12}^n, \alpha_{12}^n, \alpha_{22}^n)$, s^n moves to s^{n+1} according to the following

deterministic function

$$g(s^n,a^n,w^n)=\begin{cases}(0,0)\text{, if } i_1^n+C\left(\frac{\alpha_{11}^n p_1}{1+\alpha_{21}^n p_2^n}\right)>r_1,\\ \text{otherwise}\\ (k_1^n+1,i_1^n+C\left(\frac{\alpha_{11}^n p_1}{1+\alpha_{21}^n p_2^n}\right))\text{, if } k_1^n<N_1-1,\\ (N_1,0)\text{, if } k_1^n=N_1-1,\\ (1,i_1^n+C\left(\frac{\alpha_{11}^n p_1}{1+\alpha_{21}^n p_2^n}\right))\text{, if } k_1^n=N_1,\end{cases} \tag{46}$$

Equation (10) can be rewritten using the function g as

$$\begin{aligned}Q(B|s,a) &= \mathbb{P}\left(g(s^n,a^n,w^n)\in B|s^n=s,a^n=a\right),\\ &= \int_W \mathbb{1}_B\left(g(s,a,w)\right)f_W(w)dw,\end{aligned}$$

where, $\prod_{i,j\in\{1,2\}} f_{\alpha_{ij}}(\alpha_{ij})d\alpha_{ij}$. Due to equation (46) from a state of the form $s=(k_1^n,i_1^n)$ the only accessible states are $(0,0)$ (it corresponds to a successful decoding) or (k_1^n+1,i_1^{n+1}) (corresponding to a decoding failure).

$$Q(B|s,a) \;=\; \int_{\mathbb{S}} \mathbb{1}_B\left(s'\right)Q(ds'|s,a)$$

where $Q(ds'|s,a)$ is expressed as

$$Q(ds'|s,a)=p_0(i_1,p_2)\delta_0(dk')\delta_0(di_1')+\delta_{k_1+1}(dk')\mathbb{1}_{i_1'\geq i_1}\left(i_1'\right)f_{i_1'}(i_1'|i_1,p_2)di_1', \tag{47}$$

where $p_0(i_1,p_2)\;=\;\mathbb{P}\left(i_1+\log_2(1+\frac{\alpha_{11}p_1}{1+\alpha_{12}p_2})>r_1|i_1,\;p_2\right)$ and $f_{i_1'}(i_1'|i_1,p_2)$ is the conditional probability density function of the random variable i_1' defined as $i_1'=i_1+C\left(\frac{\alpha_{11}p_1}{1+\alpha_{21}p_2}\right)$.

If we consider now the case where $k_1=N_1-1$, after the same kind of calculation, we obtain

$$Q(ds'|s,a)=p_0\left(i_1,p_2\right)\delta_0(dk')\delta_0(di_1')+\left(1-p_0\left(i_1,p_2\right)\right)\delta_{N_1}(dk')\delta_0(di_1') \tag{48}$$

If we consider now the last case, $k_1=N_1$, we obtain

$$Q(ds'|s,a)=p_0(0,p_2)\delta_0(dk')\delta_0(di_1')+\delta_1(dk')\mathbb{1}_{i_1'\geq 0}\left(i_1'\right)f_{i_1'}(i_1'|0,p_2)di_1' \tag{49}$$

Note that we logically find here that the expression of the evolution from a state $(0,0)$ is the same as the one from a state $(N_1,0)$. This is consistent with the fact that those two classes of states correspond to the start of the transmission of a new packet.

B. Boundedness of the one-step reward functions and of the long-term reward function

By definition of the functions $R_1(s,a)$ and $R_2(s,a)$ are bounded functions for every $(s,a) \in \mathbb{K}$. For every initial distribution ν_0 and every policy $\pi \in \Pi$, $\eta_1(\nu_0,\pi)$ and $\eta_2(\nu_0,\pi)$ are then also bounded.

C. Lipschitz continuity of the one-step reward functions and of the transition kernel

We will now give some Lipschitz continuity properties. The spaces $\mathbb{S}$ and $\mathbb{A}$ are considered as subspaces of $\mathbb{R}^2$ endowed with the sup norm $||\cdot||_\infty$. We then compare two states $s = (k_1, i_1) \in \mathbb{S}$ and $s' = (k'_1, i'_1) \in \mathbb{S}$ as follows,

$$||s - s'||_\infty = \max(|k_1 - k'_1|, |i_1 - i'_1|). \tag{50}$$

Similarly, we compare two actions $a = (p_2, r_2) \in \mathbb{A}$ and $a' = (p'_2, r'_2) \in \mathbb{A}$ as follows

$$||a - a'||_\infty = \max(|p_2 - p'_2|, |r_2 - r'_2|). \tag{51}$$

We then consider $\mathbb{K}$ as a subset of $\mathbb{R}^4$ endowed with the sup norm and the distance between $k = (s,a) \in \mathbb{K}$ and $k' = (s',a') \in \mathbb{K}$ is given as follows

$$||k - k'||_\infty = \max(||s - s'||_\infty, ||a - a'||_\infty). \tag{52}$$

We finally give a distance between two probability measures on $\mathbb{S}$, p_1 and p_2 as follows

$$d(p_1,p_2) = ||p_1 - p_2||_{TV}, \tag{53}$$

$$= 2 \sup_{B \in \mathcal{B}(\mathbb{S})} |p_1(B) - p_2(B)|. \tag{54}$$

$||\cdot||_{TV}$ is called the total variation norm. For any finite signed measure on $\mathbb{S}$, m, $||\cdot||_{TV}$ is defined as follows

$$||m||_{TV} = \sup_{B \in \mathcal{B}(\mathbb{S})} m(B) - \inf_{B \in \mathcal{B}(\mathbb{S})} m(B). \tag{55}$$

Property 1 (Lipschitz continuity properties)**.** *The functions $k \mapsto R_1(k)$ and $k \mapsto R_2(k)$ and $k \mapsto Q(\cdot|k)$ are Lipschitz continuous functions of k for all $k \in \mathbb{K}$. That is, for all $k = (s,a) \in \mathbb{K}$ and $k' = (s',a') \in \mathbb{K}$ we have that there exists three positive scalars K_1, K_2 and K_Q such that*

$$|R_1(k) - R_1(k')| \leq K_1 ||k - k'||_\infty, \tag{56}$$

$$|R_2(k) - R_2(k')| \leq K_2 ||k - k'||_\infty \text{ and} \tag{57}$$

$$||Q(\cdot|k) - Q(\cdot|k')||_{TV} \leq K_Q ||k - k'||_\infty. \tag{58}$$

The two first properties after a direct application of their definition. The proof of the third one is tedious and then is omitted. Note that they imply that R_1 and R_2 are continuous functions. Property 1 also implies that Q is *strongly continuous*, that is, for every measurable bounded function u, the function $k \mapsto \int_{\mathbb{S}} u(s)Q(ds|k)$, is continuous and bounded on $\mathbb{K}$.

APPENDIX B
USEFUL DEFINITIONS AND THEOREMS

A. *On Tightness and relative compactness*

Tightness and *relative compactness* are two notions related to the study of the convergence of probability measures. Their definition is given as follows.

Definition 2 ([23] p. 186, Definition E.5)**.** *Let $\mathcal{P}$ a family of probability measures on $\mathbb{S}$.*

1) *$\mathcal{P}$ is said to be tight if and only if for every $\epsilon > 0$ there exists a compact set $S_1 \subset \mathbb{S}$ such that $\forall m \in \mathcal{P} : m(S_1) > 1 - \epsilon$.*
2) *$\mathcal{P}$ is said to be relatively compact if and only if for every every sequence in $\mathcal{P}$ contains a convergent subsequence, that is, for every sequence $\{m_n\}$ in $\mathcal{P}$ there is a subsequence $\{m_{n_i}\}$ and a probability measure m on $\mathbb{S}$ such that $m_{n_i} \overset{w.}{\to} m$.*

In this definition, the notation $m_n \overset{w.}{\to} m$ refers to the weak convergence of the sequence m_n to the measure m, that is for every continuous and bounded function v on $\mathbb{S}$,

$$\int v dm_n \underset{n\to\infty}{\to} \int v dm.$$

A link between tightness and relative compactness is done by the Prohorov's theorem given as follows.

Theorem 7 ([23], Theorem E.6 p.186)**.** *Let $\mathcal{P}$ be a family of probability measures on a metric space $\mathbb{S}$.*

1) *If $\mathcal{P}$ is tight, then it is relatively compact.*
2) *If $\mathbb{S}$ is separable and complete. If $\mathcal{P}$ is relatively compact, then it is tight.*

B. Disintegration of measure

We give a disintegration-of-measure result given in [22] or [23].

Proposition 1. *Every measure m on $\mathbb{S} \times \mathbb{A}$, for every $B \in \mathcal{B}(\mathbb{S})$ and every $C \in \mathcal{B}(A)$, there exists a stochastic kernel ψ on $\mathbb{S}$ given $\mathbb{A}$ such that*

$$m(B \times C) = \int_B \psi(C|s)\hat{m}(ds), \tag{59}$$

where $\hat{m}(B) = m(B \times \mathbb{A}),\ \forall B \in \mathcal{B}(\mathbb{S})$.

This disintegration will be denoted $m = \psi\hat{m}$.

References

[1] J. Mitola III and Jr. Maguire, G.Q., "Cognitive radio: making software radios more personal," *Personal Communications, IEEE*, vol. 6, no. 4, pp. 13 –18, aug 1999.

[2] "Report of the Spectrum Efficiency Working Group.," Tech. Rep., FCC, Nov. 2002.

[3] S. Haykin, "Cognitive radio: brain-empowered wireless communications," *IEEE Journal on Selected Areas in Communications*, vol. 23, no. 2, pp. 201–220, 2005.

[4] V. Asghari and S. Aissa, "Adaptive rate and power transmission in spectrum-sharing systems," *IEEE Transactions on Wireless Communications*, vol. 9, no. 10, pp. 3272–3280, 2010.

[5] A. Bagayoko, I. Fijalkow, and P. Tortelier, "Power control of spectrum-sharing in fading environment with partial channel state information," *IEEE Transactions on Signal Processing*, vol. 59, no. 5, pp. 2244–2256, 2011.

[6] Rui Zhang, "Optimal power control over fading cognitive radio channel by exploiting primary user csi," in *Global Telecommunications Conference, 2008. IEEE GLOBECOM 2008. IEEE*, 30 2008-dec. 4 2008, pp. 1 –5.

[7] Rui Zhang, "On active learning and supervised transmission of spectrum sharing based cognitive radios by exploiting hidden primary radio feedback," *IEEE Transactions on Communications*, vol. 58, no. 10, pp. 2960–2970, 2010.

[8] K. Eswaran, M. Gastpar, and K. Ramchandran, "Bits through arqs: Spectrum sharing with a primary packet system," in *Proceedings of the 2007 IEEE International Symposium on Information Theory (ISIT 2007)*, Nice, France, June 2007.

[9] R. A. Tannious and A. Nosratinia, "Cognitive radio protocols based on exploiting hybrid arq retransmissions," *IEEE Transactions on Wireless Communications*, vol. 9, no. 9, pp. 2833–2841, 2010.

[10] Romain Tajan, Charly Poulliat, and Inbar Fijalkow, "Opportunistic secondary spectrum sharing protocols for primary implementing an IR type Hybrid-ARQ protocol (regular paper)," in *IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP), Kyoto, 25/03/12-30/03/12*, http://www.ieee.org/, mars 2012, p. (electronic medium), IEEE.

[11] Marco Levorato, Urbashi Mitra, and Michele Zorzi, "Cognitive interference management in retransmission-based wireless networks," *CoRR*, vol. abs/1004.0542, pp. –, 2010.

[12] Eitan Altman, *Constrained Markov decision processes*, Stochastic modeling. Chapman & Hall/CRC, 1999.

[13] Frederick J Beutler and Keith W Ross, "Optimal policies for controlled markov chains with a constraint," *Journal of Mathematical Analysis and Applications*, vol. 112, no. 1, pp. 236–252, 1985.

[14] Martin L. Puterman, *Markov decision processes: discrete stochastic dynamic programming*, Wiley series in probability and statistics. Wiley-Interscience, 1994.

[15] Onésimo Hernández-Lerma, Juan González-Hernández, and Raquiel R. López-Martínez, “Constrained average cost markov control processes in borel spaces,” *SIAM J. Control Optim.*, vol. 42, no. 2, pp. 442–468, Feb. 2003.

[16] S Sesia, G Caire, and G Vivier, “Incremental redundancy hybrid arq schemes based on low-density parity-check codes,” *IEEE Transactions on Communications*, vol. 52, no. 8, pp. 1311–1321, 2004.

[17] J.F. Cheng, “Coding performance of hybrid arq schemes,” *IEEE Transactions on Communications*, vol. 54, no. 6, pp. 1017–1029, 2006.

[18] Romain Tajan, Charly Poulliat, and Inbar Fijalkow, “Interference Management for Cognitive Radio Systems Exploiting Primary IR-HARQ: a Constrained Markov Decision Process approach (regular paper),” in *IEEE Annual Asilomar Conference on Signals, Systems, and Computers (ASILOMAR), Asilomar, USA, 04/11/2012-07/11/2012*. novembre 2012, p. (electronic medium), IEEE.

[19] M. Zorzi and R. R. Rao, “On the use of renewal theory in the analysis of arq protocols,” *IEEE Transactions on Communications*, vol. 44, no. 9, pp. 1077–1081, 1996.

[20] G Caire and D Tuninetti, “The throughput of hybrid-arq protocols for the gaussian collision channel,” *IEEE Transactions on Information Theory*, vol. 47, no. 5, pp. 1971–1988, 2001.

[21] O. Hernández-Lerma, *Adaptive Markov control processes*, Applied mathematical sciences. Springer-Verlag, 1989.

[22] D.P. Bertsekas and S.E. Shreve, *Stochastic optimal control: the discrete time case*, Mathematics in science and engineering. Academic Press, 1978.

[23] Onésimo Hernández-Lerma and Jean Bernard Lasserre, *Discrete-time Markov control processes: basic optimality criteria*, vol. 30 of *Applications of mathematics*, Springer, 1996.

[24] E.J. Anderson and P. Nash, *Linear programming in infinite-dimensional spaces: theory and applications*, Wiley-Interscience series in discrete mathematics and optimization. Wiley, 1987.

[25] Chee-Seng Chow, *Multigrid algorithms and complexity results for discrete-time stochastic control and related fixed-point problems*, Ph.D. thesis, Massachusetts Institute of Technology, 1989.

[26] A.K. Karmokar, D.V. Djonin, and V.K. Bhargava, “Pomdp-based coding rate adaptation for type-i hybrid arq systems over fading channels with memory,” *Wireless Communications, IEEE Transactions on*, vol. 5, no. 12, pp. 3512 –3523, december 2006.